

RAW SEQUENCE LISTING

ERROR REPORT

BIOTECHNOLOGY
SYSTEMS
BRANCH

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following CRF diskette:

Application Serial Number: 09/510,314

Art Unit / Team No.: 01PE

Date Processed by STIC: 3/10/2000

THE ATTACHED PRINTOUT EXPLAINS THE ERRORS DETECTED.

PLEASE BE SURE TO FORWARD THIS INFORMATION TO THE APPLICANTS BY EITHER:

1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANTS ALONG WITH A NOTICE TO COMPLY or,

2) CALLING APPLICANTS AND FAXING THEM A COPY OF THE PRINTOUT WITH A NOTICE TO COMPLY

THIS WILL INSURE THAT THE NEXT SUBMISSION RECEIVED FROM THEM WILL BE ERROR FREE.

IF YOU HAVE ANY FURTHER QUESTIONS, PLEASE CALL:

MARK SPENCER 703-308-4212

Raw Sequence Listing Error Summary

ERROR DETECTED SUGGESTED CORRECTION

SERIAL NUMBER: 09/570,314

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 Wrapped Nucleics The number/text at the end of each line "wrapped" down to the next line.
This may occur if your file was retrieved in a word processor after creating it.
Please adjust your right margin to .3, as this will prevent "wrapping".
- 2 Wrapped Aminos The amino acid number/text at the end of each line "wrapped" down to the next line.
This may occur if your file was retrieved in a word processor after creating it.
Please adjust your right margin to .3, as this will prevent "wrapping".
- 3 Incorrect Line Length The rules require that a line not exceed 72 characters in length. This includes spaces.
- 4 Misaligned Amino Acid Numbering The numbering under each 5th amino acid is misaligned. This may be caused by the use of tabs between the numbering. It is recommended to delete any tabs and use spacing between the numbers.
- 5 Non-ASCII This file was not saved in ASCII (DOS) text, as required by the Sequence Rules.
Please ensure your subsequent submission is saved in ASCII text so that it can be processed.
- 6 Variable Length Sequence(s) 3 (maybe more) contain n's or Xaa's which represented more than one residue.
As per the rules, each n or Xaa can only represent a single residue.
Please present the maximum number of each residue having variable length and indicate in the (ix) feature section that some may be missing.
- 7 PatentIn ver. 2.0 "bug" A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequence(s) . Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence.
- 8 Skipped Sequences (OLD RULES) Sequence(s) missing. If intentional, please use the following format for each skipped sequence:
(2) INFORMATION FOR SEQ ID NO:X:
(I) SEQUENCE CHARACTERISTICS:(Do not insert any headings under "SEQUENCE CHARACTERISTICS")
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:X:
This sequence is intentionally skipped

Please also adjust the "(iii) NUMBER OF SEQUENCES:" response to include the skipped sequence(s).
- 9 Skipped Sequences (NEW RULES) Sequence(s) missing. If intentional, please use the following format for each skipped sequence.
<210> sequence id number
<400> sequence id number
000
- 10 Use of n's or Xaa's (NEW RULES) Use of n's and/or Xaa's have been detected in the Sequence Listing.
Use of <220> to <223> is MANDATORY if n's or Xaa's are present.
In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
- 11 Use of <213>Organism (NEW RULES) Sequence(s) are missing this mandatory field or its response.
- 12 Use of <220>Feature (NEW RULES) Sequence(s) are missing the <220>Feature and associated headings.
Use of <220> to <223> is MANDATORY if <213>ORGANISM is "Artificial" or "Unknown"
Please explain source of genetic material in <220> to <223> section.
(See "Federal Register," 6/01/98, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of new Rules)
- 13 PatentIn ver. 2.0 "bug" Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other means to copy file to floppy disk.

PAGE: 1

**RAW SEQUENCE LISTING
PATENT APPLICATION US/09/510,314**

DATE: 03/10/2000
TIME: 09:51:22

Input Set: I510314.RAW

**This Raw Listing contains the General Information
Section and up to first 5 pages.**

PP.514

```

1 <110> APPLICANT: Zhou, Ming-Ming
2 Aggarnal, Aneel K
3 <120> TITLE OF INVENTION: METHODS OF IDENTIFYING MODULATORS OF BROMODOMAINS
4 <130> FILE REFERENCE: 2459-1-003
5 <140> CURRENT APPLICATION NUMBER: US/09/510,314
6 <141> CURRENT FILING DATE: 2000-02-22
7 <160> NUMBER OF SEQ ID NOS: 44
8 <170> SOFTWARE: PatentIn Ver. 2.0
9 <210> SEQ ID NO 1
10 <211> LENGTH: 3014
11 <212> TYPE: DNA
12 <213> ORGANISM: Homo sapiens
13 <400> SEQUENCE: 1
14 gggccgcgt cgacgcggaa aagaggcgt ggggggcctc ccagcgctgg cagacaccgt 60
15 gaggctggca gcccgcggca cgcacaccta gtccgcagtc ccgagaaca tgtccgcagc 120
16 cagggcgcgg agcagagtcc cggcaggag aaccaaggga gggcgtgtc tgtggcggcg 180
17 gcggcagcgg cagcggagcc gctagtcccc tccctcctgg gggagcagct gccgcgcgtc 240
18 cccgcgcgc caccaccatc agcgcgcggg gcccgccag agcagcggg gcgagcggcg 300
19 cgctaggggg agggcggggg cggggagggg gttggcggaa gggggcggga gggcgtgggg 360
20 ggagggtctc gctctccga ctaccagac cggaggagaa ccctggggc ggcggcggcg 420
21 cctgacactc ggccctctt gccgtgtcc gggcgccat gtccgaggct ggccgggcgc 480
22 ggccggcgg ctgcggggca ggagccggg caggggcccgg gcccggggcg ctgcggggcgc 540
23 agcctgcggc gctccgcgc ggcgcgcgc aggctcccc ctgcggcgt gccgcggggg 600
24 gctcggcgc ctgcgggtccg gcgcggcag tggctgcagc gggcacggcc gaaggaccgg 660
25 gagggcgtgg ctcgccgcga atcgcgtga agaaaagcga actacgtcc gctccgcggg 720
26 ccaagaaact ggagaaactc ggagtgtact ccgcctgca gggcggaggag tctttaaat 780
27 gtaatggctg gaaaaaccct aacccctac ccactcccc cagagccgac ctgcagcaa 840
28 taattgtcag tctaacagaa tcctgtcga gtttagcca tgccctagct gctcatgttt 900
29 cccacctgga gaatgtgtca gaggaagaaa tgaacagact cctggaaata gtattggatg 960
30 tggaaatatct ctttacctgt gtccacaagg aagaagatgc agataccaaa caagtttatt 1020
31 tctatctatt taagctttt agaaaacta ttttacaaag agaaaaacct gtggttgaag 1080
32 gctcttttggaa aaagaaaaccc ccatttggaa aaccttagcat tgaacagggt gtgaataact 1140
33 ttgtgcagta caaatttagt cacctgccag caaaagaaaag gcaaacaata gttgagttgg 1200
34 caaaaatgtt cctaaaccgc atcaactatt ggcatttggaa ggcacccatct caacgaagac 1260
35 tgcgatctcc caatgtatgat atttctggat acaaagagaa ctacacaagg tggctgttt 1320
36 actgcaacgt gccacagttc tgcgcacagtc tacctcggtt cggaaaccaca caggtgttt 1380
37 ggagaacatt gcttcgtcg gtcttcaactt ttatggcgtt acaactcctg gaacaagcaa 1440
38 gacaggaaaa agataaaactg cctcttggaa aacgaactct aatcctcaact catttcccaa 1500
39 aatttctgtc catgcttagaa gaagaagttt atagtcaaaa ctctccatc tggatcagg 1560
40 attttcttc agcttctcc agaaccagcc agctaggcat ccaaaccatg atcaatccac 1620
41 ctccctgtggc tggacaattt tcatacaattt caacccatc ttcccttggag cagccaaacg 1680
42 cagggagcag cagtcctgcc tgcaaaaccccttctggact tgaggcaaac ccaggagaaa 1740
43 agagggaaaat gactgattct catgttctgg aggaggccaa gaaaccccgaa gttatggggg 1800
44 atattccat ggaattaatc aacgaggatc tgcgttccat cacggaccctt gcagcaatgc 1860

```

Does Not Comply
Corrected Diskette Needed

Input Set: I510314.RAW

45 ttggaccaga gaccaatttt ctgtcagcac actcgccag ggatgaggcg gcaagggttgg 1920
 46 aagagcgcag gggtaattt gaatttcacg tggtggcaa ttcctcaac cagaaaccaa 1980
 47 acaagaagat cctgatgtgg ctggggcc tacagaacgt tttctccac cagctcccc 2040
 48 gaatgc当地 agaatacatc acacggctcg tcttgaccc gaaacacaaa acccttgctt 2100
 49 taattaaaga tggcggtt attgggttga tctgttccg tatgttcca tctcaaggat 2160
 50 tcacagagat tgtctctgt gctgtaacct caaatgagca agtcaaggc tatggaacac 2220
 51 acctgatgaa tcatttgaaa gaatatcaca taaagcatga catcctgaac ttccatcacat 2280
 52 atgcagatga atatgcattt ggtacttta agaaacaggg tttctccaaa gaaattaaaa 2340
 53 tacctaaaac caaatatgtt ggcttatatca aggattatga aggagccact ttaatggat 2400
 54 gtgagctaa tccacggatc ccgtacacag aatttctgt catcattaaa aagcagaagg 2460
 55 agataattaa aaaactgatt gaaagaaaaac aggccacaaat tcgaaaagtt taccctggac 2520
 56 tttcatgttt taaagatgga gttcgacaga ttcctataga aagcaggctg gaaaccgagt gaaaagaga aaagtaaaga 2580
 57 agacaggctg gaaaccgagt gaaaagaga aaagtaaaga gcccagagac cctgaccagc 2640
 58 ttacagcac gctcaagagc atcctccagc aggtgaagag ccatcaaagc gcttggccct 2700
 59 tcatggaacc tgtgaagaga acagaagctc caggatatta tgaagttata aggtccccca 2760
 60 tgatctgaa aaccatgagt gaacgcctca agaatacgta ctacgtgtct aagaaattat 2820
 61 tcatggcaga cttacagcga gtcttacca attgcaaaga gtacaacgcc gctgagatg 2880
 62 aatactacaa atgtccaat atcctggaga aattcttctt cagtaaaatt aaggaagctg 2940
 63 gattaattga caagtgattt ttttcccccc tctgcttctt agaaactcac caagcagtgt 3000
 64 gcctaaagca aggt 3014
 65 <210> SEQ ID NO 2
 66 <211> LENGTH: 832
 67 <212> TYPE: PRT
 68 <213> ORGANISM: Homo sapiens
 69 <400> SEQUENCE: 2
 70 Met Ser Glu Ala Gly Gly Ala Gly Pro Gly Gly Cys Gly Ala Gly Ala
 71 1 5 10 15
 72 Gly Ala Gly Ala Gly Pro Gly Ala Leu Pro Pro Gln Pro Ala Ala Leu
 73 20 25 30
 74 Pro Pro Ala Pro Pro Gln Gly Ser Pro Cys Ala Ala Ala Gly Gly
 75 35 40 45
 76 Ser Gly Ala Cys Gly Pro Ala Thr Ala Val Ala Ala Ala Gly Thr Ala
 77 50 55 60
 78 Glu Gly Pro Gly Gly Gly Ser Ala Arg Ile Ala Val Lys Lys Ala
 79 65 70 75 80
 80 Gln Leu Arg Ser Ala Pro Arg Ala Lys Lys Leu Glu Lys Leu Gly Val
 81 85 90 95
 82 Tyr Ser Ala Cys Lys Ala Glu Glu Ser Cys Lys Cys Asn Gly Trp Lys
 83 100 105 110
 84 Asn Pro Asn Pro Ser Pro Thr Pro Pro Arg Ala Asp Leu Gln Gln Ile
 85 115 120 125
 86 Ile Val Ser Leu Thr Glu Ser Cys Arg Ser Cys Ser His Ala Leu Ala
 87 130 135 140
 88 Ala His Val Ser His Leu Glu Asn Val Ser Glu Glu Glu Met Asn Arg
 89 145 150 155 160
 90 Leu Leu Gly Ile Val Leu Asp Val Glu Tyr Leu Phe Thr Cys Val His
 91 165 170 175
 92 Lys Glu Glu Asp Ala Asp Thr Lys Gln Val Tyr Phe Tyr Leu Phe Lys
 93 180 185 190
 94 Leu Leu Arg Lys Ser Ile Leu Gln Arg Gly Lys Pro Val Val Glu Gly

PAGE: 3

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/510,314DATE: 03/10/2000
TIME: 09:51:22

Input Set: I510314.RAW

95 195 200 205
96 Ser Leu Glu Lys Lys Pro Pro Phe Glu Lys Pro Ser Ile Glu Gln Gly
97 210 215 220
98 Val Asn Asn Phe Val Gln Tyr Lys Phe Ser His Leu Pro Ala Lys Glu
99 225 230 235 240
100 Arg Gln Thr Ile Val Glu Leu Ala Lys Met Phe Leu Asn Arg Ile Asn
101 245 250 255
102 Tyr Trp His Leu Glu Ala Pro Ser Gln Arg Arg Leu Arg Ser Pro Asn
103 260 265 270
104 Asp Asp Ile Ser Gly Tyr Lys Glu Asn Tyr Thr Arg Trp Leu Cys Tyr
105 275 280 285
106 Cys Asn Val Pro Gln Phe Cys Asp Ser Leu Pro Arg Tyr Glu Thr Thr
107 290 295 300
108 Gln Val Phe Gly Arg Thr Leu Leu Arg Ser Val Phe Thr Val Met Arg
109 305 310 315 320
110 Arg Gln Leu Leu Glu Gln Ala Arg Gln Glu Lys Asp Lys Leu Pro Leu
111 325 330 335
112 Glu Lys Arg Thr Leu Ile Leu Thr His Phe Pro Lys Phe Leu Ser Met
113 340 345 350
114 Leu Glu Glu Glu Val Tyr Ser Gln Asn Ser Pro Ile Trp Asp Gln Asp
115 355 360 365
116 Phe Leu Ser Ala Ser Ser Arg Thr Ser Gln Leu Gly Ile Gln Thr Val
117 370 375 380
118 Ile Asn Pro Pro Pro Val Ala Gly Thr Ile Ser Tyr Asn Ser Thr Ser
119 385 390 395 400
120 Ser Ser Leu Glu Gln Pro Asn Ala Gly Ser Ser Ser Pro Ala Cys Lys
121 405 410 415
122 Ala Ser Ser Gly Leu Glu Ala Asn Pro Gly Glu Lys Arg Lys Met Thr
123 420 425 430
124 Asp Ser His Val Leu Glu Glu Ala Lys Lys Pro Arg Val Met Gly Asp
125 435 440 445
126 Ile Pro Met Glu Leu Ile Asn Glu Val Met Ser Thr Ile Thr Asp Pro
127 450 455 460
128 Ala Ala Met Leu Gly Pro Glu Thr Asn Phe Leu Ser Ala His Ser Ala
129 465 470 475 480
130 Arg Asp Glu Ala Ala Arg Leu Glu Glu Arg Arg Gly Val Ile Glu Phe
131 485 490 495
132 His Val Val Gly Asn Ser Leu Asn Gln Lys Pro Asn Lys Lys Ile Leu
133 500 505 510
134 Met Trp Leu Val Gly Leu Gln Asn Val Phe Ser His Gln Leu Pro Arg
135 515 520 525
136 Met Pro Lys Glu Tyr Ile Thr Arg Leu Val Phe Asp Pro Lys His Lys
137 530 535 540
138 Thr Leu Ala Leu Ile Lys Asp Gly Arg Val Ile Gly Gly Ile Cys Phe
139 545 550 555 560
140 Arg Met Phe Pro Ser Gln Gly Phe Thr Glu Ile Val Phe Cys Ala Val
141 565 570 575
142 Thr Ser Asn Glu Gln Val Lys Gly Tyr Gly Thr His Leu Met Asn His
143 580 585 590
144 Leu Lys Glu Tyr His Ile Lys His Asp Ile Leu Asn Phe Leu Thr Tyr

PAGE: 4

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/510,314DATE: 03/10/2000
TIME: 09:51:22

Input Set: I510314.RAW

145	595	600	605
146	Ala Asp Glu Tyr Ala Ile Gly Phe Lys Lys Gln Gly Phe Ser Lys		
147	610	615	620
148	Glu Ile Lys Ile Pro Lys Thr Lys Tyr Val Gly Tyr Ile Lys Asp Tyr		
149	625	630	635
150	Glu Gly Ala Thr Leu Met Gly Cys Glu Leu Asn Pro Arg Ile Pro Tyr		640
151	645	650	655
152	Thr Glu Phe Ser Val Ile Ile Lys Lys Gln Lys Glu Ile Ile Lys Lys		
153	660	665	670
154	Leu Ile Glu Arg Lys Gln Ala Gln Ile Arg Lys Val Tyr Pro Gly Leu		
155	675	680	685
156	Ser Cys Phe Lys Asp Gly Val Arg Gln Ile Pro Ile Glu Ser Ile Pro		
157	690	695	700
158	Gly Ile Arg Glu Thr Gly Trp Lys Pro Ser Gly Lys Glu Lys Ser Lys		
159	705	710	715
160	Glu Pro Arg Asp Pro Asp Gln Leu Tyr Ser Thr Leu Lys Ser Ile Leu		720
161	725	730	735
162	Gln Gln Val Lys Ser His Gln Ser Ala Trp Pro Phe Met Glu Pro Val		
163	740	745	750
164	Lys Arg Thr Glu Ala Pro Gly Tyr Tyr Glu Val Ile Arg Phe Pro Met		
165	755	760	765
166	Asp Leu Lys Thr Met Ser Glu Arg Leu Lys Asn Arg Tyr Tyr Val Ser		
167	770	775	780
168	Lys Lys Leu Phe Met Ala Asp Leu Gln Arg Val Phe Thr Asn Cys Lys		
169	785	790	795
170	Glu Tyr Asn Ala Ala Glu Ser Glu Tyr Tyr Lys Cys Ala Asn Ile Leu		800
171	805	810	815
172	Glu Lys Phe Phe Ser Lys Ile Lys Glu Ala Gly Leu Ile Asp Lys		
173	820	825	830

174 <210> SEQ ID NO 3
 175 <211> LENGTH: 12
 176 <212> TYPE: PRT
 177 <213> ORGANISM: Artificial Sequence
 178 <220> FEATURE:
 179 <223> OTHER INFORMATION: Description of Artificial Sequence: peptide
 180 <220> FEATURE:
 181 <221> NAME/KEY: VARIANT
 182 <222> LOCATION: (2) variable length not permitted - see item 6 on Env
 183 <223> OTHER INFORMATION: It represents 2 to 3 undesignated amino acids.
 184 They can be any amino acids.
 185 <220> FEATURE:
 186 <221> NAME/KEY: VARIANT
 187 <222> LOCATION: (4)
 188 <223> OTHER INFORMATION: It represents 5 to 8 undesignated amino acids.
 189 They can be any amino acids.
 190 <220> FEATURE:
 191 <221> NAME/KEY: VARIANT
 192 <222> LOCATION: (6)
 193 <223> OTHER INFORMATION: It represents one undesignated amino acid. It can
 194 be any amino acid.

(see catab position)
 of item 12 on Env
 summary sheet
 (see source of genetic material)

same
 env

PAGE: 5

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/510,314DATE: 03/10/2000
TIME: 09:51:22

Input Set: I510314.RAW

195 <220> FEATURE:
 196 <221> NAME/KEY: VARIANT *Name even*
 197 <222> LOCATION: (9)
 198 <223> OTHER INFORMATION: It represents 5 undesignated amino acids. They can
 199 be any amino acids.

200 <220> FEATURE:
 201 <221> NAME/KEY: VARIANT
 202 <222> LOCATION: (5)
 203 <223> OTHER INFORMATION: It can be any amino acid from the group of: P, K,
 204 or H.

205 <220> FEATURE:
 206 <221> NAME/KEY: VARIANT
 207 <222> LOCATION: (8)
 208 <223> OTHER INFORMATION: It can be any amino acid from the group of: Y, F,
 209 or H.

210 <220> FEATURE:
 211 <221> NAME/KEY: VARIANT
 212 <222> LOCATION: (11)
 213 <223> OTHER INFORMATION: It can be any amino acid from the group of: M, I,
 214 or V.

215 <400> SEQUENCE: 3
 W--> 216 Phe Xaa Pro Xaa Xaa Xaa Tyr Xaa Xaa Pro Xaa Asp
 217 1 5 10

218 <210> SEQ ID NO 4
 219 <211> LENGTH: 12
 220 <212> TYPE: PRT
 221 <213> ORGANISM: Artificial Sequence
 222 <220> FEATURE:
 223 <223> OTHER INFORMATION: Description of Artificial Sequence: peptide

224 <220> FEATURE:
 225 <221> NAME/KEY: SITE
 226 <222> LOCATION: (6)
 227 <223> OTHER INFORMATION: It is acetyl-lysine.

228 <400> SEQUENCE: 4
 W--OK 229 Ile Ser Tyr Gly Arg Xaa Lys Arg Arg Gln Arg Arg
 230 1 5 10

231 <210> SEQ ID NO 5
 232 <211> LENGTH: 14
 233 <212> TYPE: PRT
 234 <213> ORGANISM: Artificial Sequence
 235 <220> FEATURE:
 236 <223> OTHER INFORMATION: Description of Artificial Sequence: peptide

237 <220> FEATURE:
 238 <221> NAME/KEY: SITE
 239 <222> LOCATION: (8)
 240 <223> OTHER INFORMATION: It is acetyl-lysine.

241 <400> SEQUENCE: 5
 W--OK 242 Ala Arg Lys Ser Thr Gly Gly Xaa Ala Pro Arg Lys Gln Leu
 243 1 5 10

<210> SEQ ID NO 6

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

P/I

PAGE: 6

VERIFICATION SUMMARY
PATENT APPLICATION US/09/510,314

DATE: 03/10/2000
TIME: 09:51:22

Input Set: I510314.RAW

Line ? Error/Warning

216 W "N" or "Xaa" used: Feature required
229 W "N" or "Xaa" used: Feature required
242 W "N" or "Xaa" used: Feature required
255 W "N" or "Xaa" used: Feature required
1009 W "N" or "Xaa" used: Feature required

Original Text

Phe Xaa Pro Xaa Xaa Xaa Tyr Xaa Xaa Pro X
Ile Ser Tyr Gly Arg Xaa Lys Arg Arg Gln A
Ala Arg Lys Ser Thr Gly Gly Xaa Ala Pro A
Gln Ser Thr Ser Arg His Lys Xaa Leu Met P
Xaa Phe Xaa Pro Xaa Xaa Tyr Xaa Xaa P